Analysis of spatial correlation in predictive models of forest variables that use LiDAR auxiliary information.

F. Mauro

Oregon State University, College of Forestry, Forest Engineering Resources and Management Department. 053 Peavy Hall, Corvallis, OR 97331, USA E-mail: francisco.mauro@oregonstate.edu

V.J. Monleon

US Forest Service, Pacific Northwest Research Station, Corvallis Forestry Sciences Laboratory 3200 SW Jefferson Way, Corvallis, OR 97331, USA E-mail: vjmonleon@fs.fed.us

H. Temesgen

Oregon State University. College of Forestry, Forest Engineering Resources and Management Department. 233 Peavy Hall, Corvallis, OR 97331, USA E-mail: temesgen.hailemariam@oregonstate.edu

L.A. Ruiz

Geo-Environmental Cartography and Remote Sensing Group, Department of Cartographic Engineering, Geodesy and Photogrammetry, Universitat Politècnica de València, ESP E-mail: laruiz@cgf.upv.es

Abstract

Accounting for spatial correlation of LiDAR model errors can improve the precision of model-based estimators. To estimate spatial correlation, sample designs that provide close observations are needed, but their implementation might be prohibitively expensive. To quantify the gains obtained by accounting for the spatial correlation of model errors, we examined: 1) the spatial correlation patterns of residuals from LiDAR linear models developed to predict volume, total and stem biomass per hectare, quadratic mean diameter (QMD), basal area, mean and dominant height, and stand density; 2) the impact of field plot size on the spatial correlation patterns in a stand-wise managed Mediterranean forest in central Spain.

For all variables, the correlation range of model residuals consistently increased with plot radius and was always below 60 m except for stand density, where it reached 85 m. Except for QMD, correlation ranges of model residuals were between 1.06 and 8.16 times shorter than those observed for the raw variables. Based on the relatively short correlation ranges observed when the LiDAR metrics were used as predictors, the assumption of independent errors in many forest management inventories seems to be reasonable appropriate and appropriate in practice.

Keywords: Spatial correlation, LiDAR, forest inventory, linear models, spatial models.

1 Introduction

The use of remotely sensed auxiliary information from airborne laser scanners (ALS), in combination with the area based approach (ABA), has been an active area of research during the last two decades (Næsset 1997, Magnussen et al. 1999, Næsset and Bjerknes 2001, Andersen et al. 2005, González-Ferreiro et al. 2012), and has been implemented in operational forest inventories for more than a decade (Næsset 2002). Under the ABA, the study area is covered by a grid (i.e. a compact tessellation with non-overlapping units) containing auxiliary information for each grid cell. These grid cells are the population elements or units, so that the grid implicitly defines a pseudo sampling frame. The variables of interest and the auxiliary variables are measured in a sample of field plots and the relationship between those sets of variables is modeled. Finally, the models are used to predict the variables of interest on each grid cell. The ABA in combination with model-based estimation methods has had a prominent role in LiDAR assisted forest inventories and will be the focus of our study.

The spatial correlation of model errors has been frequently ignored in operational forest inventories assisted with remotely sensed auxiliary information, thus assuming that model errors are independent. However, assuming independence may result in loss of predictive power and incorrect variance estimators. For a spatial linear model, best linear unbiased prediction (BLUP, or kriging in the geostatistical literature) incorporates the spatial correlation to minimize the mean squared error of the prediction for unsampled locations and block averages (Cressie 1993 p. 119-167). The improvement of the prediction depends on the strength of the spatial correlation, and is greatest for the grid cells closest to the observed plots and negligible for grid cells located beyond the spatial correlation range. Assuming independence when the errors are actually correlated may result in unrealistic and typically low estimators of uncertainty (Breidenbach et al. 2016). For these reasons, there has been an increasing interest on studying the impact of the spatial correlation and how it can be incorporated in forests inventories assisted with spatially explicit auxiliary information (McRoberts, 2006; McRoberts et al., 2007; Breidenbach et al., 2008; Magnussen et al., 2009; Ver Hoef

and Temesgen, 2013; Temesgen and Ver Hoef, 2014; Finley et al., 2014; Magnussen et al., 2016a, Magnussen et al., 2016b).

To detect and model the spatial correlation of model errors, the sampling design needs to ensure the existence of pairs of observations at distances where this correlation is still present. Some studies found that the observed range of the spatial correlation of model residuals was larger than the minimum distance between plots, reaching distances of several kilometers (Magnussen et al. 2009, Ver Hoef and Temesgen 2013, Finley et al. 2014). However, in forest inventory applications, especially when using LiDAR or photogrammetric point clouds, it seems to be more common to observe residual spatial correlation that vanishes at distances from 10 m to 200 m (Breidenbach. et al. 2008, Finley et al. 2014, Breidenbach et al. 2016). These distances are typically shorter than the minimum separation between field observations e.g. (Woods et al. 2011, Rahlf et al. 2014, Mauro et al. 2016). Plots at close distance are also needed because the shape of the semivariogram at short distances is of the greatest importance for spatial prediction (Cressie 1993 p 134). The need of observations of the spatial correlation at short distance and the potentially short range of such correlation raise important questions about the design of survey protocols. Zimmerman (2006) found that sampling designs that use plots uniformly distributed throughout the study area are optimal for spatial prediction in unsampled locations when the spatial correlation parameters are known, while clustered plots are best suited for estimating the spatial correlation parameters. In practice both, spatial correlation parameters are not known and prediction at unsampled locations are needed. For this case, Zimmerman (2006) found that the best designs are those in which clusters of nearby sample plots are uniformly distributed throughout the study area.

Insights about the potential importance of the spatial correlation of model errors could be obtained from previous studies where kriging was applied to raw forest variables, without models incorporating auxiliary information (e.g., Gunnarsson et al. 1998). Based on the spatial nature of LiDAR metrics, the range of the spatial correlation should be shorter for the model residuals than for the raw variables of interest, so information from previous studies might be considered as an upper

bound for the correlation range of model residuals. However, the most desirable approach to study spatial correlation is to rely on field plots close enough to directly estimate it, which may require special sampling designs.

An additional consideration when studying spatial correlation derives from the fact that field plots and grid cells are not point, but area units. The spatial correlation has always been estimated assuming that the distance between units is the distance between their centroids (Breidenbach. et al., 2008; Magnussen et al., 2009; Finley et al., 2014; Magnussen et al., 2016 a, Magnussen et al., 2016 b). An inherent consequence of having a support area is that, as the distance between units decreases, the support areas of field plots and population grid cells can partially overlap (Figure 1). While this would not be an issue if the population is partitioned into a grid of non-overlapping units (grid cells) and a sample taken from those units, the reality of LiDAR supported ABA inventory is that the population grid and the field plot sample are misaligned for multiple reasons (i.e. positioning errors or when separation between plots is not a multiple of the grid cell size) and, therefore, field plots and grid cells partially overlap (Figure 1). Plot overlap induces correlation, since overlapping area is measured by both plots. However, estimation of the empirical autocorrelation function is the same whether there is overlap or not, simply because this is the correlation between a variable measured in two plots at a given distance between their centers, regardless of its cause. Once this empirical function is estimated, models may account for differences between overlapping and non-overlapping portions of the autocorrelation function, if warranted by the data. Because of misalignment, partial overlap between the grid that partitions the population and the field plots is a reality, so knowing the correlation at distances where the overlap is still present would help to improve the prediction for grid cells that share area with the field plots.

Field plot size has important consequences for both fieldwork cost and the ABA workflow and may also influence the spatial correlation. Larger plots usually contain more trees and thus require more measurements, which may increase the costs. However, larger plots may be less sensitive to positioning errors (Gobakken and Næsset 2009), they have less edge relative to plot area and may

result in greater precision, as the variance of the estimators is smaller in larger plots (Ruiz et al. 2014). Plot size may also affect the estimates of spatial correlation and inform the need to account for the spatial correlation of model residuals. However, to the best of our knowledge, no study has analyzed such interaction in a LiDAR assisted inventory context.

Estimation of the spatial correlation is the first and most difficult step for both spatial prediction and estimating its root mean squared error (RMSE) and is the focus of this study. Once the spatial correlation function is estimated it can be incorporated into the standard equations for spatial prediction (i.e., kriging) and variance and RMSE estimation (Cressie 1993, p. 119-167 and Searle et al. 1992, p. 272). Due to the difficulty of obtaining appropriate data, studies that examine the spatial correlation of residuals for models that use LiDAR or photogrammetric point clouds as auxiliary information are limited (Breidenbach. et al. 2008, Finley et al. 2014, Rahlf et al. 2014, Breidenbach et al. 2016). The number of variables analyzed in the literature is also very small, limited to volume and stand table data, and no previous study has considered variables such as dominant height, biomass, stand density or basal area. Thus, the objectives of this study are to:

1. Examine the spatial correlation of residuals of predictive models that use LiDAR as auxiliary information and to compare the range of the residuals to the spatial correlation ranges for the raw variables of interest. We focus on analyzing the spatial correlation at short distances, which are the most relevant for spatial prediction.
2. Examine the interaction between plot size and spatial correlation.

We examined a set of variables that can be considered as a representative sample of the type of variables used in forest management inventories. The variables examined are volume (V, m^3/ha), total biomass (B_{tot}, kg/ha), stem biomass (B_{stem}, kg/ha), quadratic mean diameter (QMD, cm), basal area (G, m^2/ha), mean tree height (H_m, m), dominant height (H_o, m) and stand density (N, stems/ha). With the exception of volume, the spatial correlation of model errors for these variables has not been examined before. Plot radii ranged from 7.5 m to 12.5 m.

2 Material and methods

2.1 Study area and areas of interest (AOI) hierarchy

The study area is a 4000 ha forest located in "La Serranía de Cuenca", central Spain, described in Ruiz et al. (2014). Approximately 5% of the area is not forested (less than 10% canopy cover (FAO 2012)). European black pine (*Pinus nigra* Arn.) and scots pine (*Pinus sylvestris* L.) are the main species and appear mixed in different proportions. Other conifers such as Spanish juniper (*Juniperus thurifera* L.) and maritime pine (*Pinus pinaster* Ait.), and hardwoods such as holm oak (*Quercus ilex* L.) and Portuguese oak (*Quercus faginea*. Lam) appear scattered over the study area. Slopes are steep and the configuration of the hydrological network, with a main river crossing the study area from north to south and several seasonal tributaries running east or west to join the main stream, result in a patch of areas with clearly differentiated slopes and orientations.

The study area is a stand-wise managed forest area, managed using shelter wood methods with a 120 years rotation period and a 20 years regeneration period. The study area contains a total of 55 delineated stands, ranging in area from 28.3 ha to 75.9 ha, which were grouped based on their similarity into 13 management units (MU) subject to similar treatments. The area of the MUs ranges from 30.6 ha to 392.3 ha.

2.2 LiDAR data

LiDAR data were collected in November 2008 using an Optech ALTM-1225 operating at 25 kHz and a maximum scanning angle of ± 18°. The average LiDAR point density was 11.4 points/m^2. Point density was not homogeneous due to irregular overlap of the scanning stripes, so the LiDAR point cloud was thinned to an homogeneous density of 4 points/m^2 using the software LAStools (Isenburg 2013). Ground points were filtered from the LiDAR point cloud and used to obtain a digital terrain model of 0.5 x 0.5 m grid cell size which was used to normalize the LiDAR point cloud. A visual inspection of the DTM and of the normalized point cloud ensured that these products were free of

spikes and outliers. All these processes were performed using FUSION (McGaughey 2014).

2.3 Field data collection

A total of 85, 25 m radius field plots were measured in December 2008. The radius of the plots used in LiDAR based forest inventories typically ranges from 9 m to 12.5 m (Ruiz et al. 2014), so the area of the plots in this study was 4 to 7.72 times larger than that of commonly used field plots. Plots were located on the nodes of a 500 m regular grid. Field crews navigated to the preselected plot centers using a navigation grade Global Positioning System (GPS) using C\A code. The coordinates of each tree with diameter at breast height (*DBH*) larger than 7 cm were obtained using compass and measuring tape. Based on previous experience, the expected accuracy of the positioning of trees relative to the plot center was approximately 0.5 m. For each tree, *DBH* was measured using a caliper, and height (*H*) using a Hagölf Vertex III hypsometer. Tree volume was computed using species specific regional equations developed by the Spanish National Forest Inventory (NFI) using *DBH* and *H* as predictors. Total tree and stem biomass were computed using species-specific models developed by Montero et al. (2005) using *DBH* as the only predictor.

Positioning errors of navigation grade GPS devices can frequently exceed 5 m and should be corrected to ensure a precise co-registration with the LiDAR data. For each plot, trees were first positioned using their coordinates relative to the plot center and overlaid on the orthophoto and on the LiDAR point cloud resulting from the filtering of the ground points. Then, an experienced photo interpreter manually adjusted the location of the plot, relying on the identification of at least seven different trees in both the digital canopy height model (DCMH), the orthophoto of the study area and the ground point cloud. Tree stem locations were identified as maxima in the DCHM and gaps in the ground point cloud derived from LiDAR. All trees in a plot were manually translated and rotated as a block until most isolated and easy to identify trees overlapped with the stem locations identified from the LiDAR image. Some trees were moved independently in each plot when their position was identified on the ground point cloud and on the orthophoto. These trees were less than a 0.5% of the

total. The average displacement of the plot center was 1.13 m, the standard deviation of the displacement was 1.72 m and the maximum displacement was 9.14 m.

2.4 Model fitting and spatial correlation assessment

For each variable of interest, we fit linear spatial models where the mean of the distribution was a function of LiDAR metrics (i.e. percentiles, moments, means, minima and maxima of LiDAR elevations as well as cover parameters, such as percentages of returns above different height thresholds (McGaughey, 2014)). We assessed the residuals to determine whether the variance was constant or, if not, we modelled it as function of one of the LiDAR predictors. The general model fitting strategy consisted on starting with simple models and adding complexity if needed. First, we selected the LiDAR metrics to include in the model, then considered a weighting schema to account for heteroscedasticity, added a random effect for the management unit and, finally, modeled the residual spatial autocorrelation as a function of the distance between subplot centers.

2.4.1 Computation of subplot level values and auxiliary information

For each 25 m radius plot, groups of subplots of radius between 7.5 and 12.5 m, in 0.5 m increments contained within the larger plot were created. Each group of subplots was obtained by first defining a subplot concentric to the 25 m radius plot. New subplots were defined by moving outwards the central subplot in steps of 0.5 m following E-W, SE-NW, S-N and SW-NE directions, until the edge of the subplots were tangent to the border of the 25 m radius plot. The number of subplots in each 25 m plot, the total number of subplots and the maximum distances between subplot centers are shown in Table 1. Note that the number of steps and the maximum distance between subplot centers (max_distance(radius)) is different for each radius (Table 1) and equals 50 m minus two times the subplot radius. For each subplot, the variables of interest were calculated and expanded to a per hectare basis when appropriate. Similarly, the full set of 30 LiDAR metrics were computed for each subplot using FUSION (McGaughey 2014).

2.4.2 Non-spatial models

First, we fitted linear fixed effects models of the variables of interest as a function of the LiDAR metrics. Because of the very large number of potential predictor variables, we selected parsimonious models by first choosing the fixed effects using the R package leaps (Lumley 2009) based on the 12.5 m radius subplot only. The maximum number of predictors was set to 3 independent variables per model. We obtained the best 5 models in terms of adjusted coefficient of determination when considering 1, 2 and 3 auxiliary variables for a total of 15 models for each variable of interest. These models were denoted as $m_{0,vrbl,l}$ where $vrbl$ is a sub-index to denote the variable of interest, and subscript l = 1, 2,..., 15 indicates the candidate model.

Typically, the variance of the model residuals was not constant, so we fit a new set of 15 models, $m_{1,vrbl,l}$ accounting for heteroscedasticity using the R package nlme (Pinheiro et al. 2015). For each of the 15 models selected previously, the standard deviation of the errors was assumed to be proportional to a power of the predictor most correlated with the variable of interest, mcp^{η}. The errors e_{ij} for the j^{th} grid cell in the i^{th} management unit can then be written as $e_{ij} = \sigma_{e0} mcp_l^{\eta} \epsilon_{ij}$ where ϵ_{ij} is the standardized error, distributed as $\epsilon_{ij} \sim N(0,1)$, so that $V(e_{ij})^{1/2} = \sigma_{e0} mcp_{l,i,j}^{\eta}$, and η is a model parameter. Models $m_{0,vrbl,l}$and $m_{1,vrbl,l}$ can both be written as

[1] $$y_{ij} = \boldsymbol{\beta} \boldsymbol{x}_{ij} + e_{ij}$$

where y_{ij} and $\boldsymbol{x}_{ij}$ are the variable of interest and the vector of auxiliary variables for the j^{th} grid cell in the i^{th} management unit, respectively, and $\boldsymbol{\beta}$ is a vector of parameters. The model error is e_{ij}, with variance $V(e_{ij}) = \sigma_{0e}^2$ for $m_{0,vrbl,l}$ and $V(e_{ij}) = \sigma_{0e}^2 mcp_{l,i,j}^{2\eta}$ for $m_{1,vrbl,l}$. Note that the model $m_{0,vrbl,l}$ is nested within $m_{1,vrbl,l}$ when $\eta = 0$. Thus, we compared those two models using a likelihood-ratio test and selected the simplest model, $m_{0,vrbl,l}$, when including the heteroscedasticity did not improved the model fit significantly, and $m_{1,vrbl,l}$ otherwise (Pinheiro and Bates 2000 p 84).

We added a management unit random effect to the best models selected in the previous step, resulting in the model:

[2] $$y_{ij} = \boldsymbol{\beta} \boldsymbol{x}_{ij} + v_i + e_{ij}$$

Here v_i is the random effect of the i^{th} management unit, which are assumed to be independent and identically distributed variables with mean 0 and variance $V(v_i) = \sigma_v^2$. These models were denoted as $m_{2,vrbl,l}$ and both $m_{0,vrbl,l}$ and $m_{1,vrbl,l}$ are nested within $m_{2,vrbl,l}$, when $\sigma_v^2 = 0$. Therefore, the significance of this random effect was tested using a likelihood-ratio test (Pinheiro and Bates 2000 p 84). Selected models were denoted by $m^{*}_{vrbl,l}$. Both Pearson´s standardized residuals and the management unit random effects were graphically assessed and the best behaved candidate model, in terms of normality and heteroscedasticity, was selected and denoted by m^{**}_{vrbl}.

Finally, for the other subplot radii we kept constant the fixed effects selected for the 12.5 m radius plot model, m^{**}_{vrbl} , but tested for the need to account for heteroscedasticity and MU random effects as described above. The resulting models were denoted as $m^{**}_{vrbl,rad}$, where the sub-index rad indexes the subplot radius.

2.4.3 Spatial correlation assessment

To analyze the spatial correlation, we computed the Pearson correlation of the standardized Pearson residuals ($\hat{\epsilon}_{i,j}$) from the final model, $m^{**}_{vrbl,rad}$ for all pairs of subplots separated at distance d between 0.5 m and $max\_distance(radius)$, in 0.5 m intervals, but using only pairs of observations on the same moving line (dashed-lines Figure 1). We computed the correlation of the residuals for each subplot radius, distance, and variable of interest, which yields a directional empirical correlation function. Then, all the pairs were pooled together to compute an isotropic empirical correlation function (i.e., the correlogram, Cressie 1993 p.67). Empirical correlations at distance d are denoted hereafter as $\omega_{vrbl,rad}(d)$, where the subscripts are described in the previous section. We examined the correlation as a function of distance between subplot centers to select a suitable spatial correlation model for each variable of interest.

The covariance of the model errors for locations *b* and *c* within the same MU was expressed as

$Cov(e_b, e_c) = \sigma_b \sigma_c G(d_{b,c}, \rho, \theta)$, where σ_b and σ_c are the standard deviation of the model errors for the subplots centered at locations b and c, respectively; $G(d_{b,c}, \rho, \theta)$ (Eq. [3]) is the correlation function; $d_{b,c}$ is the Euclidean distance between those locations; and ρ, θ are parameters. We modeled the correlation function as:

$$[3] \quad G(d_{b,c}, \rho, \theta) = \theta I(d_{b,c} < 2rad)\left\{\left[acos\left(\frac{d_{b,c}}{2rad}\right)\right]\frac{2}{\pi} - \left(\frac{d_{b,c}}{\pi rad^2}\sqrt{rad^2 - \frac{{d_{b,c}}^2}{4}}\right)\right\} + (1-\theta)e^{-(\frac{d_{b,c}}{\rho})}$$

Note that $Cov(e_b, e_c)$ is a function of the location, because σ_b and σ_c may depend on mcp, which can change from point to point. On the other hand, $G(d_{b,c}, \rho, \theta)$ is the covariance, as well as the correlation function for the standardized errors ϵ_{ij}. The distribution of the standardized errors is stationary (Cressie 1993 p. 57), with mean 0 and variance 1, and the correlation only depends on the distance between plot centers. Then, modeling $G(d_{b,c}, \rho, \theta)$ is functionally equivalent to modeling the semivariogram for the standardized residuals (Cressie 1993 p. 67).

The shape of the model for $G(d_{b,c}, \rho, \theta)$ was chosen after observing the empirical correlation function. It is a mixture of two components: the first component accounts for the correlation when there is overlap between subplots (this is, when the distance between subplot centers is less than twice their radius) and the second component is a pure exponential model without nugget effect. $I(d_{b,c} < 2rad)$ is an indicator function of whether there is overlap or not and θ is the weight for the first component. The effective or practical range, denoted as φ hereafter, is defined as the distance for which the correlation descends to 0.05 (Bivand et al. 2008 p. 202), and it is a function of ρ, θ and the plot radius.

Due to the spatial nature of LiDAR metrics and their high explanatory power, one can expect a reduction of the spatial correlation of model errors when compared to the spatial correlation of the raw variables. To assess this reduction, we examined the spatial correlation of the raw variables and compared it to the spatial correlation of the residuals. As with the residuals, we modeled the spatial correlation patterns for the response variables using the correlation function in [3]. For each variable

of interest and subplot radius empirical correlations, the covariance function, the correlation function, its parameters and the effective range ($w(d), Cov(d_{b,c}, \rho, \theta), G(d_{b,c}, \rho, \theta), \rho, \theta, \varphi$) were indexed using sub-indexes $vrbl$ and rad to denote the variable of interest and the subplot radius. A super-script res or raw was added to indicate model residuals or raw variables respectively. We computed the ratios of the effective empirical correlation ranges of the residuals and raw variables $\gamma_{vrbl,rad}^{raw,res} = \frac{\varphi_{vrbl,rad}^{raw}}{\varphi_{vrbl,rad}^{res}}$ to summarize the reduction of the range of the spatial correlation when the auxiliary information is used to predict the response variable.

2.4.4 Influence of plot size in the spatial correlation of the residuals

To analyze how plot size interacts with the spatial correlation of the residuals and raw variables, we first computed the correlation coefficients between the range of the spatial correlation and the plot radius. Then, we conducted a more detailed analysis by directly examining the observed empirical correlations instead of the parameters of the spatial correlation model. In this analysis, we also excluded the distances for which there was subplot overlap. For each variable and subplot distance between 20 and 30 m we tested the effect of increasing the plot radius on the empirical correlation of model residuals using Kendal's τ test (Dalgaard 2008 p. 124).

3 Results and discussion

The exploratory analysis revealed that empirical correlation of model residuals decreased with distance between subplot centers without marked differences among directions, so that the isotropic models without nugget effect (i.e., Eq. 1) were appropriate (Figure 3), and can be used to derive kriging estimates as well as mean square error estimators (Searle et al. 1992, ch. 7, Cressie 1993, p. 119-167). For basal area and subplot radii 7.5, 8.5, 9 and 9.5 m, and for tree density and subplot radii of 9.5, 10 and 10.5 m, the right tail of the empirical correlation function for the residuals was particularly flat. This led to models with a very large ρ parameter for the exponential component and, therefore, unrealistically large values of the estimated effective ranges. This effect was especially

marked for basal area, where the computed ranges for those subplot radii were orders of magnitude longer than those observed for the remaining subplots radii for the same variables or for other variables. Those seven cases were removed from the analysis of the spatial correlation vs plot radius.

For all the raw variables, the effective spatial correlation range estimated from model [3] was always less than 200 m (Figure 4). For the residuals, the effective range was always below 60 m, with the exception of tree density and the seven cases discussed in the previous paragraph. QMD had the shortest range among the raw variables and, in addition, the LiDAR variables did not explain much of its variability. As a result, when the LiDAR variables were included, the spatial correlation range of the residuals increased slightly and $\gamma_{QMD,rad}^{raw,res}$ was close to, but less than one. Tree density, which typically shows the weakest relationship with LiDAR auxiliary information (Næsset 2002, Goerndt et al. 2010), had the longest correlation range for model residuals and the ratio $\gamma_{N,rad}^{raw,res}$ ranged from 1.06 to 3.45. After QMD, the variables with the shortest correlation ranges were volume, dominant height and mean tree height, while the remaining variables (B_{tot}, B_{stem}, G) showed correlation ranges for the residuals that were larger than those observed for QMD, H_m, H_o and V, but smaller than those observed for N. The high predictive power of LiDAR data for dominant height (Næsset 1997, Magnussen et al. 1999, Næsset and Økland 2002, Næsset and Bjerknes 2001, Goerndt et al. 2010 explains why the spatial correlation range of the residuals was 3.14 to 5.14 times shorter than that observed for the raw variable ($\gamma_{H_o,rad}^{raw,res}$) and became even shorter than the spatial correlation range of the residuals of mean height (Figure 4). For volume, the values of $\gamma_{V,rad}^{raw,res}$ were similar to or greater than those for dominant height and considerable larger than those for total and stem biomass (Figure 4 and Table A 1). This could be explained by the strong relationship between LiDAR metrics with tree height, a variable that was included in the tree volume equations but not in the biomass equations. For the biomass variables and basal area, the reduction of the spatial correlation range when the LiDAR auxiliary information was included was larger than that observed for QMD, but smaller than that observed for H_o and V. Both, B_{tot}, B_{stem} and G, are related in different ways to N and tree height, which may explain this average behavior.

The range of the spatial correlation consistently increased with plot radius for both the model residuals and the raw responses. For all variables, except basal area and tree density for which four and three radius were excluded, the increase was statistically significant for the residuals ($p<0.05$, Figure 4). To exclude the possible effect of plot overlap on the correlation, we examined the relationship between plot radius and the empirical spatial correlation only for the distances where there was no overlap. This included 19 different distances for each variable (152 pairs of variable-distance). For 136 cases (approximately 90% of the cases) the Kendall's τ coefficient was positive, indicating that the larger the radius, the larger the empirical correlation of model residuals (Figure 5). This result suggests that assuming uncorrelated errors might have a larger impact when using larger plots, which raises questions for further research about the use of different plot sizes.

Our sampling design did not allow us to examine the spatial dependence at distances larger than 25 to 35 meters depending on the subplot radius. However, for all variables and subplot radii, the empirical correlation at the maximum possible distance between subplot centers was always below 0.26 and, in most cases, did not exceed 0.1 (Table *2*). For 39% (34 out of 88) of the cases, the empirical correlation at the maximum distance was below 0.05. Therefore, the distances considered in the sample always covered more than 74% of maximum possible value of the empirical spatial correlation of model residuals. In most cases the coverage was greater than 90%, and in more than a third of the cases it rached the effective range (Table *2*). While extrapolations based on a spatial correlation model would be needed for predictions at distances larger than the maximum distances examined here, the greatest influence of the spatial correlation for both prediction and estimation of uncertainty is at the shortest distances (Cressie 1993 p 134.). In addition, as confirmed empirically for most variables, at the largest distances examined the spatial correlation almost disappeared (Table *2*) and extrapolation errors would be bounded by a very small value.

The results obtained in this study emphasize the need of sampling designs that include very close observations if the modeler aims to analyze or use the spatial correlation of model errors. For those close distances, partial overlap between plots could be a concern, but overlap between training

plots and the population grid cells occurs in real applications and, therefore, it should not be disregarded. Incorporating the influence of overlap into the spatial correlation function, as we did here, should not be particularly difficult. The design used in this study also provides distances without overlap, allowing estimation of the correlation at those distances. Even if the overlap was considered a problem, if the spatial correlation models fit well in the section without overlap the effect of plot overlap causes no harm in subsequent estimates.

The need to incorporate the spatial correlation of model errors in model based inference for forest inventory has received significant attention recently, as it may result on improved grid cell and stand level predictions and uncertainty estimators (Magnussen et al. 2016a) and help scale-up LiDAR predictions from different inventories made with different plot or grid cell sizes to a common size (Magnussen et al. 2016b). However, most studies recognize that directly estimating the spatial correlation of model errors requires costly field observations, so they proposed indirect methods to estimate this correlation. Typically, those methods overcome the lack of direct measurements by relying on strong assumptions, such as proportionality of the spatial correlation range of predictions and model errors, which cannot be empirically confirmed in the context of those studies. The spatial correlation models obtained here are empirical results that can be used directly to improve predictions or to estimate the uncertainty of the predictions. The ratios $\gamma_{vrbl,rad}^{raw,res}$ could be used to anticipate the spatial correlation of model errors if previous information about the raw forest attributes of interest, such as that provided by Gunnarsson et al. (1998), were available in a similar study area.

The demanding fieldwork needed to estimate the spatial correlation is the main reason why few studies analyze the correlation of residuals, especially from models based on LiDAR auxiliary information or photogrammetric point clouds (Breidenbach. et al. 2008, Finley et al. 2014, Rahlf et al. 2014, Breidenbach et al. 2016). Breidenbach et al. (2008) modeled volume as a function of LiDAR variables and reported spatial correlations that decrease below 0.05 for distances of 202 m of larger. The study examined pairs of observations at distances of 100, 200, 223 m and larger, so that only two

of those distances were within the spatial correlation range, which makes those results lack some robustness. Breidenbach et al. (2016) also examined volume and found correlation ranges for the residuals smaller than 100m. However, while the design of this study was more consistent and included multiple pairs of observations at close distances, it did not consider LiDAR auxiliary information. Rahlf et al. (2014) also examined volume using a sampling design with plots of 9 m radius and pairs of observations at distances as small as 20 m. These authors did not observe any spatial correlation for the residuals. Those results are in accordance with the results obtained in this study for volume where the spatial correlation ranges varied between 13.65 m and 21.34 m (Appendix A, Table A 1). Finally, Finley et al. (2014) analyzed stand tables (i.e. number of stems in predefined diameter classes) and also found spatial correlation ranges smaller than 100 m for certain diameter classes. In our study, we examined a large set of inventory variables and also found that the spatial range was relatively short, particularly for the residuals of the variables that are better explained by LiDAR metrics.

The estimated spatial correlation range of the residuals was so short in most cases that the assumption of independent residuals (e.g Woods et al. (2011), Mauro et al. (2016)) seems to be reasonably accurate. For prediction, the effect of omitting the spatial correlation in a situation like the one observed in this study can be illustrated with the following hypothetical example. If we consider a systematic design with plots on the nodes of a rectangular grid, a density of 0.1 plot/ha, a management unit of 50 ha and a grid cell size of 15 m, each management unit would contain approximately 2222 grid cells and 5 plots. If the spatial correlation of the residuals vanishes at 40 m, the errors associated to each field plot could be assumed to be independent for model fitting purposes. Incorporating the spatial structure to improve the predictive performance of the models would have very little impact compared to a model that assumes independence: it would only affect the prediction for about 35 grid cells per measured plot (i.e., grid cells that surround a plot and are closer than 40 m). In total, only predictions for around 175 grid cells out of 2222 (approximately the 8% of the total number of grid cells in the MU) would be different from those omitted the spatial

correlation.

The effect of omitting the spatial correlation on the computation of variances and mean squared errors of predictions typically results in overoptimistic uncertainty measures. Breidenbach et al. (2016) examined this effect using 9 m radius plots in a case where the range of the spatial correlation of model errors was 69 m. The omission of the spatial correlation resulted in variance estimates that were 15% smaller than those obtained when accounting for the spatial correlation. However, based on the short spatial correlation range of models errors estimated in this study, the effect of omitting the spatial correlation on uncertainty measures should presumably be small, at least for areas of interest of relatively large size. When using smaller plots, the effect of the spatial correlation might be very small, on the other hand, because of the greater range when using larger plots, modelling the spatial correlation may result in increased precision in the predictions. This problem deserves more analysis, and further studies should consider different plot sizes and forest attributes. The results obtained here provide a reference about the spatial correlation expected for a wide set of variables.

While our results are specific for the study area, for volume they are in accordance to previous experiences (Rahlf et al. 2014), and in general they suggest some guidelines for future inventories in similar areas. First, commonly used systematic sampling designs seem inappropriate for modeling spatial correlation of LiDAR model errors in forest management inventories. Even with plots overlaid on the nodes of an extremely dense grid, where distance between nodes is 100 m (1 plot/ha), estimating the spatial correlation would be difficult or even impossible for all variables. Sample designs with randomly located plots could provide pairs of observations within the correlation ranges, but high sampling efforts should be done due to the short correlation ranges observed for residuals of most variables. Based on our experience, the best alternatives for operational forest inventories that seek to analyze the spatial correlation of model errors are: 1) to include or complement the field information with at least some large plots with georeferenced tree positions, like the ones used in this study, or 2) use a plot shape consisting of clusters of subplots such as that used in some national

inventories, e.g. the U.S. Forest Inventory and Analysis (Bechtold and Patterson 2005). If the first alternative is used, it requires obtaining tree coordinates relative to the plot center as well as the absolute coordinates of the plot center. Mapping every tree can be very time consuming, but promising techniques based on terrestrial laser scanners (Liang et al. 2016) or even photogrammetric point clouds obtained using inexpensive cameras (e.g. Gatziolis et al. 2015), may provide centimetric accuracy for relative coordinates, while current GPS technology allows very precise positioning of the field plot center. For the second alternative, it may be necessary to incorporate subplots at a greater range of distances, as the actual FIA design, for example, only allows to examine two distances between subplots (36.58 m and 63.35 m). In any case, both designs would allow obtaining clusters of observations, which is the option Zimmerman (2006) recommended to optimize sample designs that account for spatial correlation of errors when both fixed effects and spatial correlation parameters are unknown.

4 Conclusions

When auxiliary information with high explanatory power is available, the assumption of uncorrelated errors implicitly accepted in a large number of LiDAR assisted forest inventory applications seems to be reasonably accurate. This study showed that the spatial range of the residuals was so short that misspecification by ignoring the spatial correlation may not have a significant effect on model predictions. However, the effect of such misspecification on uncertainty measures needs to be further examined.

In general, when LiDAR information is included the spatial correlation range of the residuals is smaller than that of the raw variables. The reduction is greatest for variables highly correlated with LiDAR. The spatial correlation ranges of both, model residuals and raw variables, increases with plot radius.

Sampling designs with clusters of plots separated by small distances are needed to study spatial correlation, as this tends to vanish at distances shorter than the minimum separation between plots

than the one employed in most LiDAR assisted inventories.

Acknowledgements

The authors wish to thank Drs. Jay Ver Hoef and Isabel Molina for their valuable comments on earlier versions of the manuscript. The US Bureau of Land Management and the Spanish Ministries of Industry, Tourism and Trade, and Science and Innovation provided financial support in the framework of the projects: *Use of LIDAR and other remote sensing data with FIA plots for mapping forest inventory in Southwest Oregon, InForest II TSI-020100-2009-815*, and *CGL2010-19591/BTE*.

5 References

Andersen, H.-E., McGaughey, R.J., and Reutebuch, S.E. 2005. Estimating forest canopy fuel parameters using LIDAR data. Remote Sens. Environ. **94**(4): 441–449.

Bechtold, W.A., and Patterson, P.L. 2005. The enhanced forest inventory and analysis program-national sampling design and estimation procedures.

Bivand, R.S., Pebesma, E.J., and Gomez-Rubio, V. 2008. Applied spatial data analysis with R. Available from http://www.asdar-book.org/.

Breidenbach., J., Kublin, E., McGaughey, R., Andersen, H.E., and Reutebuch, S. 2008. Mixed-effects models for estimating stand volume by means of small footprint airborne laser scanner data. Photogramm. J. Finl. **21**(1): 4–15.

Breidenbach, J., McRoberts, R.E., and Astrup, R. 2016. Empirical coverage of model-based variance estimators for remote sensing assisted estimation of stand-level timber volume. Remote Sens. Environ. **173**: 274–281. doi:10.1016/j.rse.2015.07.026.

Cressie, N.A.C. 1993a. Geostatistics. *In* Statistics for Spatial Data. John Wiley & Sons, Inc. pp. 27–104. Available from http://dx.doi.org/10.1002/9781119115151.ch2.

Cressie, N.A.C. 1993b. Spatial Prediction and Kriging. *In* Statistics for Spatial Data. John Wiley & Sons, Inc. pp. 105–209. Available from http://dx.doi.org/10.1002/9781119115151.ch3.

Dalgaard, P. 2008. Introductory Statistics with R. *In* 2nd edition. Springer New York, New York, NY. Available from http://link.springer.com/10.1007/978-0-387-79054-1 [accessed 31 October 2016].

FAO. 2012. Forest Resources Assessment Working Paper 180. FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS., Rome.

Finley, A.O., Banerjee, S., Weiskittel, A.R., Babcock, C., and Cook, B.D. 2014. Dynamic spatial

regression models for space-varying forest stand tables. Environmetrics **25**(8): 596–609. doi:10.1002/env.2322.

Gatziolis, D., Lienard, J.F., Vogs, A., and Strigul, N.S. 2015. 3D Tree Dimensionality Assessment Using Photogrammetry and Small Unmanned Aerial Vehicles. PLoS ONE **10**(9): e0137765. doi:10.1371/journal.pone.0137765.

Gobakken, T., and Næsset, E. 2009. Assessing effects of positioning errors and sample plot size on biophysical stand properties derived from airborne laser scanner data. Can. J. For. Res. **39**(5): 1036–1052.

Goerndt, M.E., V. J. Monleon, and H Temesgen. 2010. Relating forest attributes with area-and tree-based light detection and ranging metrics for western Oregon Western Journal of Applied Forestry 25 (3), 105-111.

González-Ferreiro, E., Diéguez-Aranda, U., and Miranda, D. 2012. Estimation of stand variables in Pinus radiata D. Don plantations using different LiDAR pulse densities. Forestry.

Gunnarsson, F., Holm, S., Holmgren, P., and Thuresson, T. 1998. On the potential of Kriging for forest management planning. Scand. J. For. Res. **13**(1–4): 237–245. doi:10.1080/02827589809382981.

Isenburg, M. 2013. LAStools - efficient tools for LiDAR processing. http://lastools.org. Available from http://rapidlasso.com.

Liang, X., Kankare, V., Hyyppä, J., Wang, Y., Kukko, A., Haggrén, H., Yu, X., Kaartinen, H., Jaakkola, A., Guan, F., Holopainen, M., and Vastaranta, M. 2016. Terrestrial laser scanning in forest inventories. Theme Issue State---Art Photogramm. Remote Sens. Spat. Inf. Sci. **115**: 63–77. doi:10.1016/j.isprsjprs.2016.01.006.

Lumley, T. 2009. leaps: regression subset selection. Available from http://CRAN.R-project.org/package=leaps.

Magnussen, S., Eggermont, P., and LaRiccia, V.N. 1999. Recovering Tree Heights from Airborne Laser Scanner Data. For. Sci. **45**: 407–422(16).

Magnussen, S., Frazer, G., and Penner, M. 2016a. Alternative mean-squared error estimators for synthetic estimators of domain means. J. Appl. Stat.: 1–24. doi:10.1080/02664763.2016.1142942.

Magnussen, S., Mandallaz, D., Lanz, A., Ginzler, C., Næsset, E., and Gobakken, T. 2016b. Scale effects in survey estimates of proportions and quantiles of per unit area attributes. For. Ecol. Manag. **364**: 122–129. doi:10.1016/j.foreco.2016.01.013.

Magnussen, S., McRoberts, R.E., and Tomppo, E.O. 2009. Model-based mean square error estimators for k-nearest neighbour predictions and applications using remotely sensed data for forest inventories. Remote Sens. Environ. **113**(3): 476–488.

Mauro, F., Molina, I., García-Abril, A., Valbuena, R., and Ayuga-Téllez, E. 2016. Remote sensing estimates and measures of uncertainty for forest variables at different aggregation levels. Environmetrics **27**(4): 225–238. doi:10.1002/env.2387.

Mc Gaughey, R.J. 2014. FUSION/LDV: Software for LIDAR Data Analysis and Visualization.

McRoberts, R.E. 2006. A model-based approach to estimating forest area. Remote Sens. Environ. **103**(1): 56–66. doi:10.1016/j.rse.2006.03.005.

McRoberts, R.E., Tomppo, E.O., Finley, A.O., and Heikkinen, J. 2007. Estimating areal means and variances of forest attributes using the k-Nearest Neighbors technique and satellite imagery. Remote Sens. Environ. **111**(4): 466–480. doi:10.1016/j.rse.2007.04.002.

Montero, G., Ruiz-Peinado, R., Muñoz, M., España. Ministerio de Educación y Ciencia, and Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria. 2005. Producción de biomasa y fijación de CO2 por los bosques españoles. Ministerio de Educación y Ciencia. Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria. Available from https://books.google.com/books?id=j4o41tVIV40C.

Næsset, E. 1997. Determination of mean tree height of forest stands using airborne laser scanner data. ISPRS J. Photogramm. Remote Sens. **52**(2): 49–56.

Næsset, E. 2002. Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. Remote Sens. Environ. **80**(1): 88–99.

Næsset, E., and Bjerknes, K.-O. 2001. Estimating tree heights and number of stems in young forest stands using airborne laser scanner data. Remote Sens. Environ. **78**(3): 328–340.

Næsset, E., and Økland, T. 2002. Estimating tree height and tree crown properties using airborne scanning laser in a boreal nature reserve. Remote Sens. Environ. **79**(1): 105–115.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., and Team, R.C. 2015. nlme: Linear and Nonlinear Mixed Effects Models.

Pinheiro, J.C., and Bates, D.M. 2000. Mixed-Effects Models in S and S-Plus. Springer.

Rahlf, J., Breidenbach, J., Solberg, S., Næsset, E., and Astrup, R. 2014. Comparison of four types of 3D data for timber volume estimation. Remote Sens. Environ. **155**: 325–333. doi:10.1016/j.rse.2014.08.036.

Ruiz, L.A., Hermosilla, T., Mauro, F., and Godino, M. 2014. Analysis of the Influence of Plot Size and LiDAR Density on Forest Structure Attribute Estimates. Forests **5**(5): 936–951. doi:10.3390/f5050936.

Searle, S.R., Casella, G., and McCulloch, C.E. 1992. Prediction of Random Variables. *In* Variance Components. John Wiley & Sons, Inc. pp. 258–289. Available from http://dx.doi.org/10.1002/9780470316856.ch7.

Temesgen, H., and Ver Hoef, J.M. 2014. Evaluation of the spatial linear model, random forest and

gradient nearest-neighbour methods for imputing potential productivity and biomass of the Pacific Northwest forests. Forestry. doi:10.1093/forestry/cpu036.

Ver Hoef, J.M., and Temesgen, H. 2013. A Comparison of the Spatial Linear Model to Nearest Neighbor (k-NN) Methods for Forestry Applications. PLoS ONE **8**(3): e59129. doi:10.1371/journal.pone.0059129.

Woods, M., Pitt, D., Penner, M., Lim, K., Nesbitt, D., Etheridge, D., and Treitz, P. 2011. Operational implementation of a LiDAR inventory in Boreal Ontario. For. Chron. **87**(04): 512–528. doi:10.5558/tfc2011-050.

Zimmerman, D.L. 2006. Optimal network design for spatial prediction, covariance parameter estimation, and empirical prediction. Environmetrics **17**(6): 635–652. doi:10.1002/env.769.

Tables

Table 1. Number of subplots, maximum distance, and number of pairs of observations at 0.5m and at max_distance used to estimate the spatial correlation parameters.

Subplot radius	Subplots/25 m plot	Total subplots	Max distance (m)	Pairs at 0.5 m	Pairs at max distance
7.5	284	24140	35	23800	
8	276	23460	34	23120	
8.5	268	22780	33	22440	
9	260	22100	32	21760	
9.5	252	21420	31	21080	
10	244	20740	30	20400	340
10.5	236	20060	29	19720	
11	228	19380	28	19040	
11.5	220	18700	27	18360	
12	212	18020	26	17680	
12.5	204	17340	25	17000	

Table 2. Empirical correlation of the model residual for each variable and subplot radius at the maximum distance. $.max\_distance(radius) = 50 - 2radius$

	Variable							
Radius(m)	V	B_{tot}	B_{stem}	G	H_m	H_o	QMD	N
7.5	0.04	0.07	0.07	0.15	-0.03	0.02	0.12	0.26
8	-0.02	0.03	0.04	0.03	-0.03	0.02	0.10	0.26
8.5	0.03	0.10	0.08	0.19	-0.03	0.02	0.11	0.26
9	0.02	0.11	0.08	0.19	0.00	0.00	0.12	0.26
9.5	0.02	0.11	0.08	0.19	0.03	0.02	0.13	0.26
10	0.01	0.11	0.08	0.19	0.06	0.02	0.13	0.25
10.5	0.00	0.12	0.08	0.20	0.09	0.04	0.12	0.24
11	0.01	0.12	0.09	0.21	0.10	0.03	0.11	0.25
11.5	0.02	0.12	0.09	0.06	0.13	0.04	0.10	0.24
12	0.03	0.05	0.07	0.06	0.14	0.05	0.10	0.24
12.5	0.06	0.03	0.04	0.07	0.16	0.05	0.10	0.24

Figure captions

Figure 1. Field plot and grid of pixels. Note the partial overlap between the field plot and the four pixels around its center.

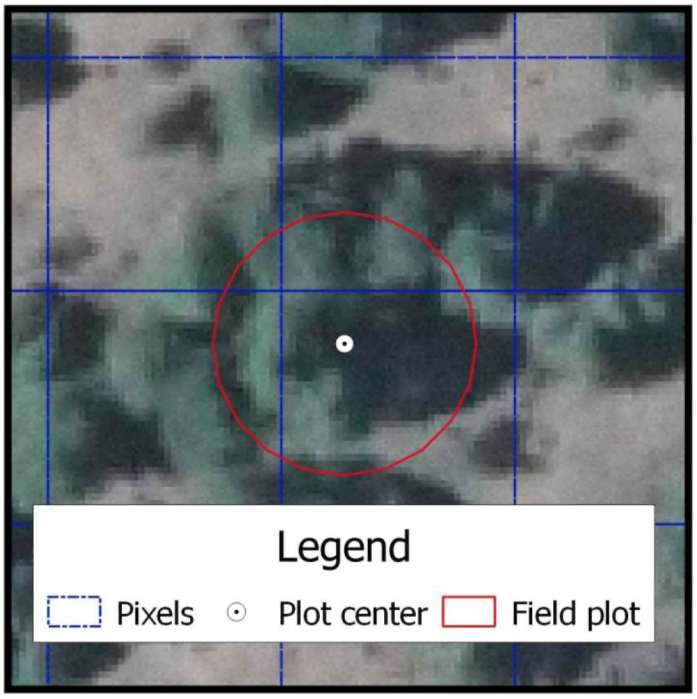
Figure 2. Example of 25 m radius plot and subplots of radius 7.5 m, 10 m and 12.5 m (for clarity, other subplots radii are omitted) moving in an East to West direction. South East to North West, South to North and South West to North East directions. Field plots were moved in 0.5 m steps and are marked with dashed lines. Trees are plotted according to their crown radius

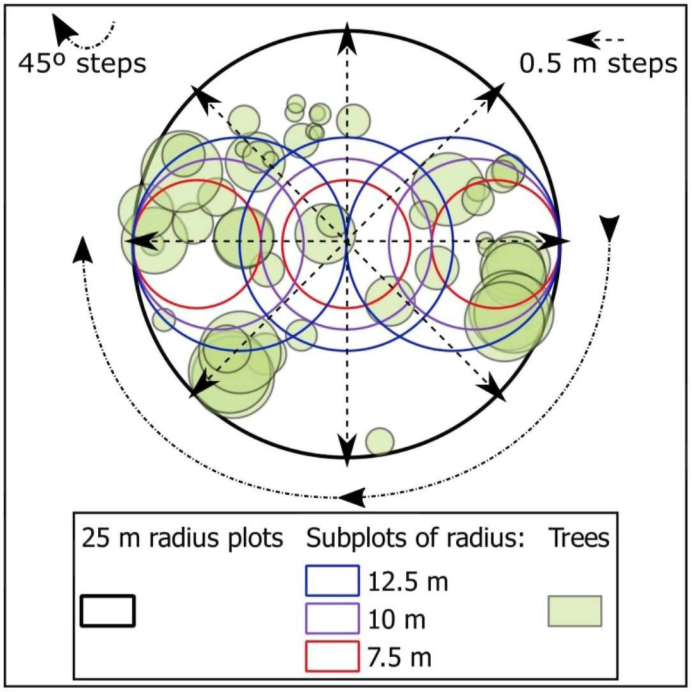
Figure 3. Spatial correlation models for the residuals from $m^{**}_{vrbl,rad}$ and for the raw variables.

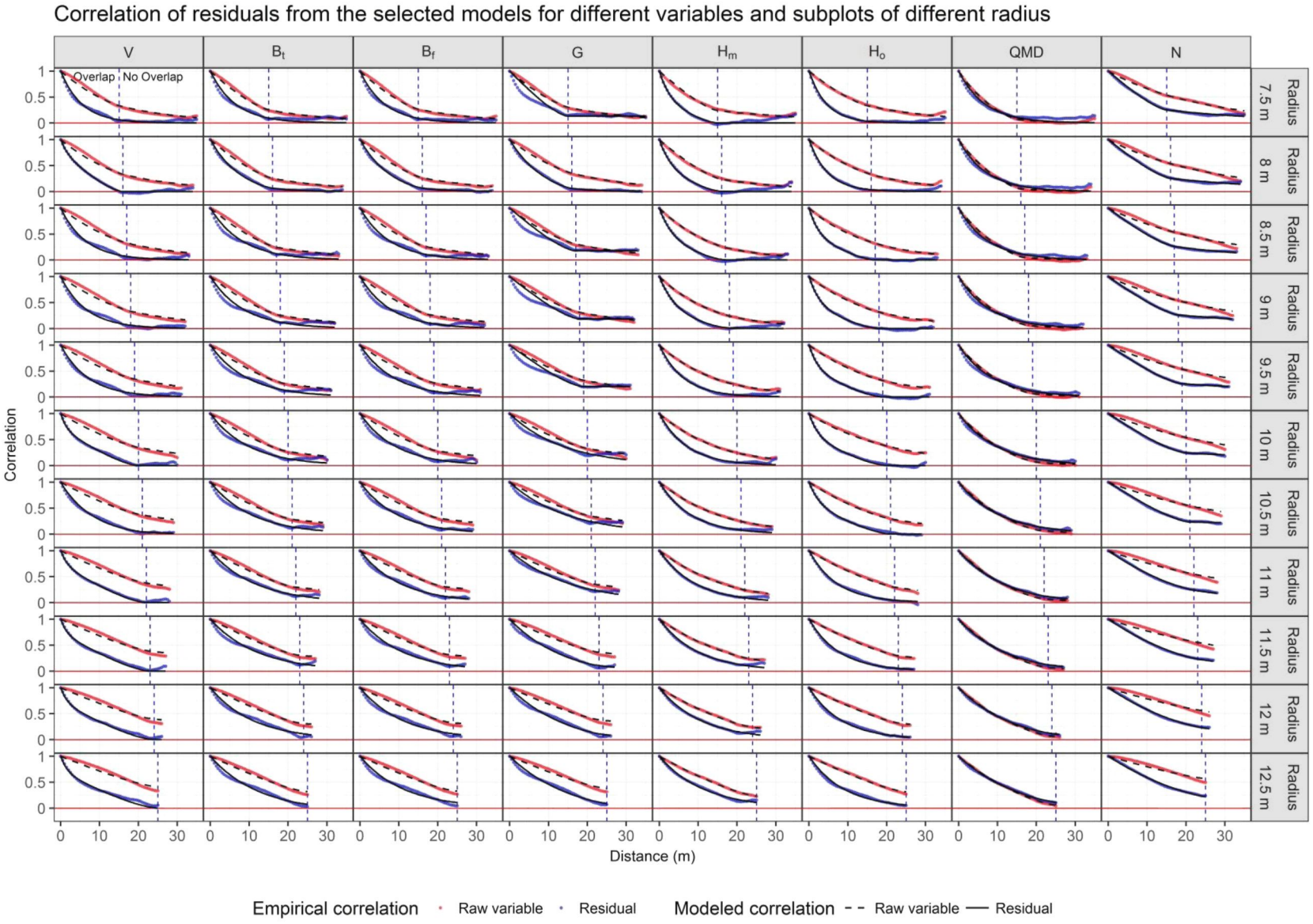
Figure 4. Distances (φ) for which correlation between pairs of observations decreases to 0.05 and parameters $\gamma_{vrbl,rad}^{raw,res}$.

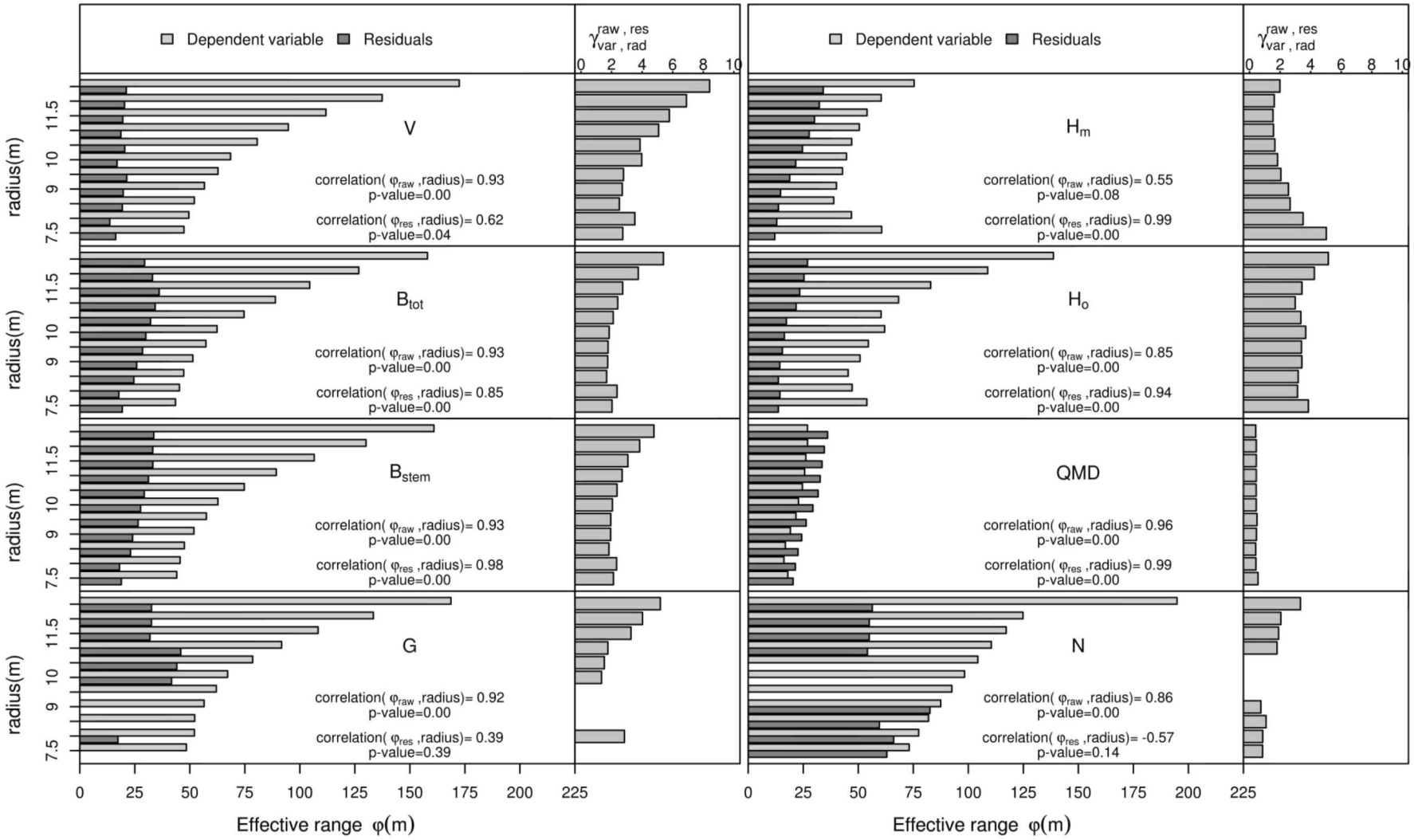
Figure 5. Results for the Kendall's τ significance test for each variable and subplot radius. Only non-overlapping plots are considered. Subfigures a,b,c and d are examples included as a graphical legend for the figure in the upper panel.

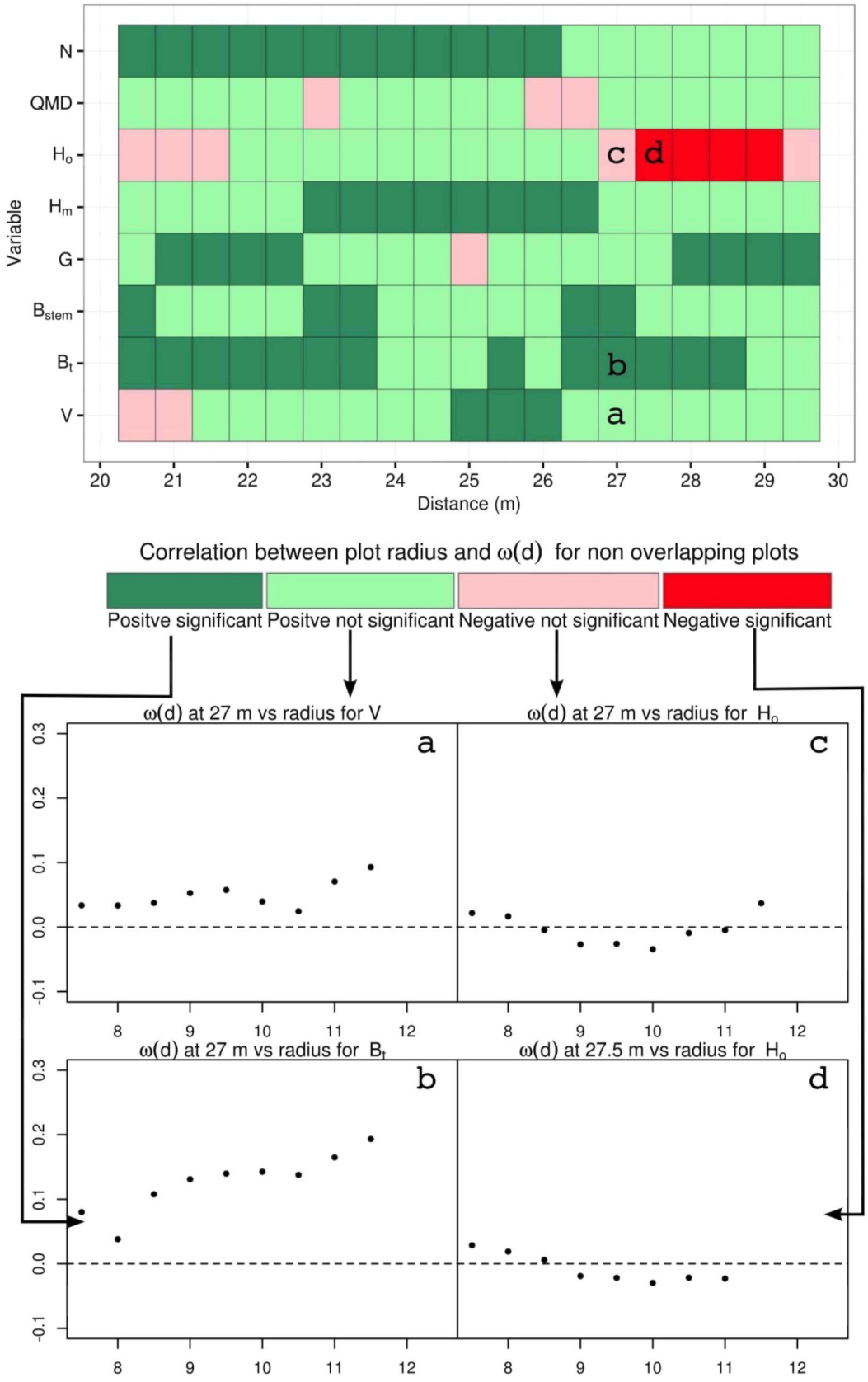
Figures




45º steps
0.5 m steps
25 m radius plots
Subplots of radius:
12.5 m
10 m
7.5 m
Trees


Correlation of residuals from the selected models for different variables and subplots of different radius
V
B_t
B_f
G
H_m
H_o
QMD
N
Overlap
No Overlap
Radius
7.5 m
8 m
8.5 m
9 m
9.5 m
10 m
10.5 m
11 m
11.5 m
12 m
12.5 m
Correlation
Distance (m)
Empirical correlation
Raw variable
Residual
Modeled correlation
Raw variable
Residual


Dependent variable
Residuals
$\gamma^{raw,\,res}_{var,\,rad}$
V
correlation(φ_{raw} ,radius)= 0.93
p-value=0.00
correlation(φ_{res} ,radius)= 0.62
p-value=0.04
B_{tot}
correlation(φ_{raw} ,radius)= 0.93
p-value=0.00
correlation(φ_{res} ,radius)= 0.85
p-value=0.00
B_{stem}
correlation(φ_{raw} ,radius)= 0.93
p-value=0.00
correlation(φ_{res} ,radius)= 0.98
p-value=0.00
G
correlation(φ_{raw} ,radius)= 0.92
p-value=0.00
correlation(φ_{res} ,radius)= 0.39
p-value=0.39
H_m
correlation(φ_{raw} ,radius)= 0.55
p-value=0.08
correlation(φ_{res} ,radius)= 0.99
p-value=0.00
H_o
correlation(φ_{raw} ,radius)= 0.85
p-value=0.00
correlation(φ_{res} ,radius)= 0.94
p-value=0.00
QMD
correlation(φ_{raw} ,radius)= 0.96
p-value=0.00
correlation(φ_{res} ,radius)= 0.99
p-value=0.00
N
correlation(φ_{raw} ,radius)= 0.86
p-value=0.00
correlation(φ_{res} ,radius)= -0.57
p-value=0.14
radius(m)
Effective range φ(m)


N
QMD
H_o
H_m
G
B_{stem}
B_t
V
Variable
c
d
b
a
20
21
22
23
24
25
26
27
28
29
30
Distance (m)
Correlation between plot radius and ω(d) for non overlapping plots
Positve significant
Positve not significant
Negative not significant
Negative significant
ω(d) at 27 m vs radius for V
ω(d) at 27 m vs radius for H_o
ω(d) at 27 m vs radius for B_t
ω(d) at 27.5 m vs radius for H_o
a
c
b
d
0.3
0.2
0.1
0.0
-0.1
8
9
10
11
12

Appendix A

Table A 1 Estimated variance and spatial correlation parameters for each variable of interest and subplots radius. The most correlated predictor (mcp) for each variable of interest is denoted with an asterisk. The parameter σ_v is the standard deviation of the management unit random effects, σ_{0e} and η are the parameters that model the variance of model errors and ρ^{res} and θ^{res} and ρ^{raw} and θ^{raw} are the parameters of the spatial correlation model for the residuals and the raw variables respectively. $\varphi^{res}_{vrbl,rad}$ and $\varphi^{dep}_{vrbl,rad}$ represent the effective spatial correlation range of residuals and raw variables and $\gamma^{raw,res}_{vrbl,rad}$ denotes the ratio $\frac{\varphi^{raw}_{vrbl,rad}}{\varphi^{res}_{vrbl,rad}}$. ElvP99, ElvP95 ElvP75, ElvP40 and ElvP01 are the 99th, 95th, 75th,40th and 1st percentile of the LiDAR returns height. Elvmean and ElvAAD are the mean LiDAR height and the average absolute deviation from the mean LiDAR height. CRR is the canopy relief ratio and, RtAbvmean , PercRtAbvmea and PercR1tAbvmea represent the total number of returns above the mean, the percentage of returns above the mean, and the percentage of first returns above the mean height respectively.

Variable of interest (Model predictors)	rad	σ_{0e}	η	σ_v	Residuals ρ^{res}	Residuals θ^{res}	Residuals φ^{res}	Raw Variables ρ^{raw}	Raw Variables θ^{raw}	Raw Variables φ^{raw}	$\gamma^{raw,res}_{vrbl,rad}$
V(m^3/ha) (Elvmean*, ElvP01, ElvP40)	7.5	14.68	0.78	9.01E-02	5.44	0.00	16.32	16.93	0.18	47.34	2.80
	8.0	33.51	--	1.19E-01	1.84	0.75	13.65	17.60	0.16	49.55	2.82
	8.5	12.68	0.82	3.40E-01	6.45	0.00	19.34	18.60	0.18	52.05	2.80
	9.0	12.05	0.83	3.72E-01	6.58	0.00	19.74	20.29	0.18	56.63	2.79
	9.5	11.59	0.81	3.57E-01	7.12	0.00	21.35	22.62	0.20	62.81	2.78
	10.0	11.12	0.81	3.73E-01	1.47	0.71	16.93	24.86	0.21	68.51	2.76
	10.5	10.71	0.81	4.42E-01	7.22	0.17	20.44	29.90	0.26	80.64	2.70
	11.0	10.22	0.81	6.08E-01	1.69	0.73	18.69	36.00	0.30	94.77	2.63
	11.5	9.77	0.81	7.67E-01	1.50	0.73	19.55	43.40	0.34	111.85	2.58
	12.0	9.26	0.83	8.67E-01	1.67	0.72	20.34	54.81	0.39	137.40	2.51
	12.5	8.98	0.80	9.18E-01	1.68	0.70	21.13	70.25	0.42	172.44	2.45
B_{tot} (kg/ha) (Elvmean*, CRR)	7.5	12217.26	0.78	4.56E-02	6.44	0.00	19.32	17.47	0.40	43.49	2.49
	8.0	24586.30	--	1.94E-01	6.51	0.24	17.75	17.98	0.38	45.29	2.52
	8.5	11224.13	0.76	9.41E-02	8.19	0.00	24.58	18.89	0.39	47.24	2.50
	9.0	10830.21	0.74	1.08E-01	8.62	0.00	25.86	20.51	0.39	51.39	2.51
	9.5	10439.33	0.73	1.39E-01	9.51	0.00	28.54	22.99	0.39	57.33	2.49
	10.0	10064.25	0.73	1.87E-01	10.01	0.00	30.02	25.11	0.40	62.37	2.48

Variable of interest (Model predictors)	rad	σ_{0e}	η	σ_v	Residuals			Raw Variables			$\gamma_{vrbl,rad}^{raw,res}$
					ρ^{res}	θ^{res}	φ^{res}	ρ^{raw}	θ^{raw}	φ^{raw}	
	10.5	9746.68	0.72	2.52E-01	10.72	0.00	32.16	30.72	0.43	74.65	2.43
	11.0	9433.28	0.71	3.53E-01	11.42	0.00	34.27	37.62	0.47	88.88	2.36
	11.5	9154.14	0.70	4.57E-01	12.03	0.00	36.09	45.02	0.49	104.49	2.32
	12.0	9322.38	0.60	5.95E-01	11.01	0.00	33.03	56.31	0.52	126.78	2.25
	12.5	9168.77	0.56	7.31E-01	10.56	0.19	29.46	72.00	0.55	158.00	2.19
B_{stem} (kg/ha) (Elvmean*, CRR)	7.5	8124.30	0.79	5.64E-02	6.29	0.00	18.87	17.32	0.36	44.08	2.55
	8.0	16910.59	--	1.99E-01	6.58	0.23	18.04	17.67	0.34	45.58	2.58
	8.5	7419.93	0.77	1.50E-01	7.68	0.00	23.05	18.54	0.35	47.52	2.56
	9.0	7144.41	0.76	1.62E-01	8.00	0.00	23.99	20.23	0.35	51.87	2.56
	9.5	6868.77	0.75	1.80E-01	8.83	0.00	26.50	22.48	0.35	57.52	2.56
	10.0	6590.74	0.75	2.18E-01	9.21	0.00	27.63	24.73	0.37	62.77	2.54
	10.5	6370.98	0.74	2.72E-01	9.77	0.00	29.30	30.20	0.41	74.75	2.48
	11.0	6142.53	0.74	3.80E-01	10.40	0.00	31.19	37.15	0.45	89.29	2.40
	11.5	5912.13	0.74	4.95E-01	11.07	0.00	33.20	45.23	0.47	106.54	2.36
	12.0	5827.23	0.70	6.16E-01	11.06	0.00	33.17	56.83	0.51	130.08	2.29
	12.5	5756.19	0.66	7.33E-01	11.21	0.00	33.62	72.04	0.53	160.93	2.23
QMD (cm) (ElvP95*, RtAbvmean, ElvP99)	7.5	7.60	0.04	5.93E-02	6.80	0.00	20.39	7.53	0.45	18.01	2.39
	8.0	8.03	-0.01	6.47E-02	7.14	0.00	21.41	7.37	0.55	16.24	2.20
	8.5	8.39	-0.05	6.68E-02	7.54	0.00	22.62	7.61	0.54	16.87	2.22
	9.0	8.53	-0.08	6.55E-02	8.10	0.00	24.29	8.08	0.47	19.15	2.37
	9.5	8.38	-0.08	6.68E-02	8.77	0.00	26.32	8.86	0.42	21.69	2.45
	10.0	8.48	-0.11	6.60E-02	9.79	0.00	29.37	9.64	0.47	22.82	2.37
	10.5	8.66	-0.13	6.34E-02	10.59	0.00	31.77	10.36	0.46	24.63	2.38
	11.0	8.79	-0.16	6.24E-02	11.31	0.10	32.69	10.99	0.48	25.64	2.33
	11.5	8.86	-0.18	6.31E-02	11.59	0.10	33.54	11.02	0.46	26.18	2.38
	12.0	8.70	-0.18	6.62E-02	12.26	0.16	34.58	11.66	0.50	26.96	2.31
	12.5	8.62	-0.20	6.69E-02	13.38	0.26	36.06	12.91	0.60	26.92	2.09
G (m2/ha) (Elvmean*, CRR)	7.5	2.82	0.74	5.65E-06	1E05	0.87	1E05	18.46	0.31	48.45	2.62
	8.0	5.16	--	1.46E-05	6.73	0.34	17.35	18.39	0.15	52.09	2.83
	8.5	2.55	0.73	1.09E-05	1E05	0.82	1E05	19.36	0.26	52.23	2.70
	9.0	2.45	0.72	1.31E-05	1E05	0.81	1E05	20.90	0.25	56.49	2.70
	9.5	2.36	0.71	1.64E-05	1E05	0.80	1E05	22.98	0.26	62.05	2.70
	10.0	2.29	0.70	2.07E-05	13.90	0.00	41.69	24.95	0.26	67.20	2.69
	10.5	2.24	0.69	2.55E-05	14.70	0.00	44.11	29.75	0.30	78.58	2.64
	11.0	2.18	0.67	3.17E-05	15.30	0.00	45.89	35.63	0.34	91.73	2.57
	11.5	2.59	0.34	3.77E-05	11.64	0.23	31.89	42.95	0.38	108.30	2.52
	12.0	2.23	0.45	4.93E-05	12.11	0.26	32.56	54.59	0.42	133.41	2.44
	12.5	2.13	0.44	5.86E-05	12.24	0.28	32.59	70.87	0.46	168.67	2.38
H_m (m) (ElvAAD*, ElvP75)	7.5	2.71	-0.04	1.54E-01	2.87	0.41	12.07	29.05	0.60	60.63	2.09
	8.0	2.64	-0.07	1.54E-01	3.22	0.41	12.97	19.90	0.47	46.91	2.36
	8.5	2.58	-0.10	1.53E-01	3.37	0.41	13.73	13.90	0.18	38.86	2.80
	9.0	2.53	-0.13	1.59E-01	4.27	0.30	14.68	13.96	0.12	40.06	2.87

Variable of interest (Model predictors)	rad	σ_{0e}	η	σ_v	Residuals			Raw Variables			$\gamma_{vrbl,rad}^{raw,res}$
					ρ^{res}	θ^{res}	φ^{res}	ρ^{raw}	θ^{raw}	φ^{raw}	
	9.5	2.48	-0.16	1.63E-01	6.29	0.00	18.86	14.80	0.10	42.84	2.89
	10.0	2.43	-0.17	1.63E-01	7.20	0.00	21.59	15.57	0.12	44.66	2.87
	10.5	2.37	-0.19	1.66E-01	8.21	0.00	24.64	16.43	0.13	47.00	2.86
	11.0	2.31	-0.19	1.76E-01	9.23	0.00	27.70	17.94	0.17	50.48	2.81
	11.5	2.24	-0.20	1.89E-01	10.04	0.00	30.12	19.48	0.20	53.99	2.77
	12.0	2.16	-0.20	2.07E-01	10.76	0.00	32.29	23.26	0.33	60.43	2.60
	12.5	2.08	-0.19	2.25E-01	11.36	0.00	34.07	32.80	0.50	75.39	2.30
H_o (m) (ElvAAD*)	7.5	2.83	-0.04	1.87E-01	4.56	0.00	13.69	20.46	0.30	53.90	2.63
	8.0	2.77	-0.08	2.02E-01	4.80	0.00	14.39	16.54	0.13	47.20	2.85
	8.5	2.72	-0.11	2.19E-01	2.38	0.48	13.68	15.14	0.00	45.42	3.00
	9.0	2.56	-0.14	2.09E-01	2.77	0.44	14.35	17.73	0.12	50.79	2.86
	9.5	2.53	-0.16	2.36E-01	2.24	0.54	15.52	19.25	0.15	54.59	2.84
	10.0	2.48	-0.18	2.54E-01	2.34	0.56	16.42	20.66	0.00	61.99	3.00
	10.5	2.44	-0.18	2.80E-01	2.59	0.58	17.34	22.22	0.24	60.38	2.72
	11.0	2.33	-0.20	2.81E-01	7.26	0.00	21.77	25.92	0.30	68.32	2.64
	11.5	2.28	-0.20	3.35E-01	7.80	0.00	23.39	33.52	0.41	82.91	2.47
	12.0	2.17	-0.21	3.41E-01	8.44	0.00	25.32	46.48	0.48	108.83	2.34
	12.5	2.11	-0.21	3.98E-01	8.98	0.00	26.94	61.90	0.53	138.67	2.24
N (stems/ha) (PercR1Abvmean*, PercRtAbvmean , ElvAAD)	7.5	7.13	0.94	1.61E+01	29.60	0.58	62.98	24.38	0.00	73.13	3.00
	8.0	6.11	0.97	2.22E+01	30.73	0.57	66.09	25.82	0.00	77.45	3.00
	8.5	5.34	1.00	3.18E+01	25.70	0.49	59.60	27.31	0.00	81.93	3.00
	9.0	4.88	1.02	3.98E+01	38.90	0.58	82.57	29.15	0.00	87.46	3.00
	9.5	4.70	1.02	4.43E+01	53.15	0.63	105.99	30.84	0.00	92.52	3.00
	10.0	4.42	1.03	5.25E+01	63.42	0.67	120.40	32.76	0.00	98.28	3.00
	10.5	4.16	1.04	6.12E+01	145.03	0.73	244.45	34.78	0.00	104.33	3.00
	11.0	3.90	1.05	7.10E+01	20.50	0.30	54.16	36.82	0.00	110.45	3.00
	11.5	3.63	1.06	8.17E+01	20.80	0.30	54.97	39.07	0.00	117.21	3.00
	12.0	3.40	1.08	9.03E+01	20.25	0.24	55.00	41.62	0.00	124.86	3.00
	12.5	3.17	1.09	1.01E+02	21.25	0.29	56.38	69.89	0.19	194.84	2.79